ORIGINAL

# DESARROLLO Y ENTRENAMIENTO DE UNA RED NEURONAL CONVOLUCIONAL PARA LA DETECCIÓN DE ESOFAGITIS EN IMÁGENES ENDOSCÓPICAS

*Development and training of a convolutional neural network for the detection of esophagitis in endoscopic images*

**Diego Martínez R, Cano de la Cruz JD, Sánchez Sánchez MI, Vázquez Pedreño LA, Jiménez Pérez M**

HOSPITAL REGIONAL UNIVERSITARIO DE MÁLAGA

## Resumen

**Introducción y objetivos:** La endoscopia digestiva ofrece una evaluación directa del tracto gastrointestinal, aunque la variabilidad entre operadores puede limitar su precisión. Este estudio se propuso desarrollar una red neuronal convolucional (CNN) basada en InceptionResNetV2, ajustada para la detección automatizada de esofagitis en imágenes endoscópicas, con el objetivo de mejorar la exactitud diagnóstica y optimizar el flujo clínico.

**Material y métodos:** Se implementó el modelo utilizando Python, Keras y TensorFlow en Google Colab Pro con GPU Nvidia A100. Partiendo de la arquitectura InceptionResNetV2 preentrenada en ImageNet, se añadieron capas densas para realizar una clasificación binaria (línea Z normal vs. esofagitis). El entrenamiento se realizó con 2000 imágenes del conjunto KVASIR (80% entrenamiento y 20% validación). La evaluación se extendió a 1164 imágenes del conjunto HyperKVASIR, excluyendo casos leves, y a 203 imágenes del Hospital Regional Universitario de Málaga.

**Resultados:** El modelo demostró altos índices de acierto, evidenciados en matrices de confusión y curvas ROC, con AUC de 0.884 para el conjunto KVASIR y 0.970 para HyperKVASIR. Se observó una mayor precisión en la detección de esofagitis avanzadas, correlacionando la severidad de la lesión con un incremento en la exactitud diagnóstica.

**Conclusiones:** El estudio destaca el potencial de las CNN en el diagnóstico asistido por IA en endoscopia. Aunque el modelo muestra alta sensibilidad en lesiones avanzadas, se requieren investigaciones adicionales para mejorar la detección en estadios incipientes y validar su aplicación en contextos clínicos heterogéneos.

*Raúl Diego Martínez*
*Hospital Regional Universitario de Málaga*
*raul.diego.martinez@outlook.com*

*Diego Martínez R, Cano de la Cruz JD, Sánchez Sánchez MI, Vázquez Pedreño LA, Jiménez Pérez M. Desarrollo y entrenamiento de una red neuronal convolucional para la detección de esofagitis en imágenes endoscopicas. RAPD 2025;48(2):48-53. DOI: 10.37352/2025482.1*

**Palabras clave**: endoscopia digestiva, redes neuronales convolucionales, esofagitis, deep learning, inteligencia artificial.

## Abstract

**Introduction and objectives:** digestive endoscopy provides a direct evaluation of the gastrointestinal tract, although inter-operator variability can limit its precision. This study aimed to develop a convolutional neural network (CNN) based on InceptionResNetV2, tailored for the automated detection of esophagitis in endoscopic images, with the objective of improving diagnostic accuracy and optimizing clinical workflow.

**Materials and methods:** the model was implemented using Python, Keras, and TensorFlow on Google Colab Pro with an Nvidia A100 GPU. Starting from the InceptionResNetV2 architecture pretrained on ImageNet, dense layers were added to perform binary classification (normal Z-line vs. esophagitis). Training was conducted using 2000 images from the KVASIR dataset (80% for training and 20% for validation). Evaluation was extended to 1164 images from the HyperKVASIR dataset, excluding mild cases, and to 203 images from the Hospital Regional Universitario de Málaga.

**Results:** the model demonstrated high accuracy, as evidenced by confusion matrices and ROC curves, with an AUC of 0.884 for the KVASIR dataset and 0.970 for HyperKVASIR. Greater precision was observed in the detection of advanced esophagitis, correlating the severity of the lesion with increased diagnostic accuracy.

**Conclusions:** the study highlights the potential of CNNs in AI-assisted diagnosis in endoscopy. Although the model shows high sensitivity in advanced lesions, additional research is required to improve detection in early stages and to validate its application in heterogeneous clinical contexts.

**Keywords**: digestive endoscopy, convolutional neural networks, esophagitis, deep learning, artificial intelligence.

## Introducción

La endoscopia digestiva permite ofrecer una evaluación directa y mínimamente invasiva del tracto gastrointestinal, la evaluación, sin embargo, sigue estando sujeta a cierta variabilidad entre operadores, lo que provoca diferencias significativas en la eficiencia de la técnica según quién la lleve a cabo.

En este contexto, la inteligencia artificial (IA) y, en particular, las redes neuronales convolucionales (CNN, por sus siglas en inglés), han emergido como herramientas prometedoras para mejorar la precisión y reproducibilidad del diagnóstico endoscópico. Estas arquitecturas de aprendizaje profundo, inspiradas en la forma en que las neuronas del cerebro humano se interconectan, están compuestas por múltiples capas de neuronas artificiales que se van entrenando para reconocer patrones a partir de datos. Durante el entrenamiento, cada capa extrae características o rasgos de nivel creciente de complejidad, ajustando constantemente los pesos de sus conexiones a fin de mejorar su capacidad de clasificación o detección. En la práctica clínica actual, la IA se ha comenzado a utilizar en la detección de pólipos colorrectales, la caracterización de lesiones gástricas y otras aplicaciones que buscan apoyar el diagnóstico endoscópico[1,2].

El presente trabajo se centra en el desarrollo de una red neuronal convolucional entrenada específicamente para la detección de esofagitis en imágenes endoscópicas. A través del uso de técnicas de procesamiento de imágenes y aprendizaje profundo, se busca mejorar la capacidad diagnóstica automatizada. Este enfoque no solo tiene el potencial de mejorar la toma de decisiones clínicas, sino también de agilizar el flujo de trabajo en entornos médicos, facilitando un diagnóstico más rápido y preciso para los pacientes.

En este artículo, se describirá el proceso de desarrollo y entrenamiento del modelo de IA, así como su validación mediante un conjunto de datos de imágenes endoscópicas. Además, se discutirán los desafíos y perspectivas futuras en la integración de estos sistemas en la práctica clínica, con el objetivo de mejorar la calidad del diagnóstico endoscópico y la atención a los pacientes con enfermedades esofágicas.

## Materia y métodos

En este proyecto se ha utilizado el lenguaje de programación Python junto con las bibliotecas Keras y TensorFlow[3] para implementar y entrenar una arquitectura de red neuronal profunda. El entorno de trabajo seleccionado fue Google Colab Pro, que proporciona acceso a potentes unidades de procesamiento gráfico (GPU), en este caso una Nvidia A100. Esto resulta fundamental para reducir significativamente los tiempos de entrenamiento y poder manejar grandes volúmenes de datos de imágenes.

La arquitectura base escogida fue InceptionResNetV2[4], desarrollada originalmente por investigadores de Google y entrenada con el conjunto de datos masivo y público de ImageNet[5]. InceptionResNetV2 combina las ventajas de las

 DOI: 10.37352/2025482.1

convoluciones de la familia Inception con la estabilidad y eficiencia de las conexiones de tipo residual, dando como resultado un modelo que mantiene un equilibrio adecuado entre precisión y velocidad de entrenamiento.

El modelo original de InceptionResNetV2, tras haberse entrenado en la clasificación de miles de categorías de ImageNet, se adaptó para nuestro caso de uso específico. Para ello, se reemplazaron las capas de salida por capas personalizadas diseñadas para realizar clasificación binaria. En particular, se añadieron tres capas densas (completamente conectadas) que culminan en una capa de salida con la activación idónea (para distinguir entre dos clases: esofagitis versus una línea Z norma (Figura 1).

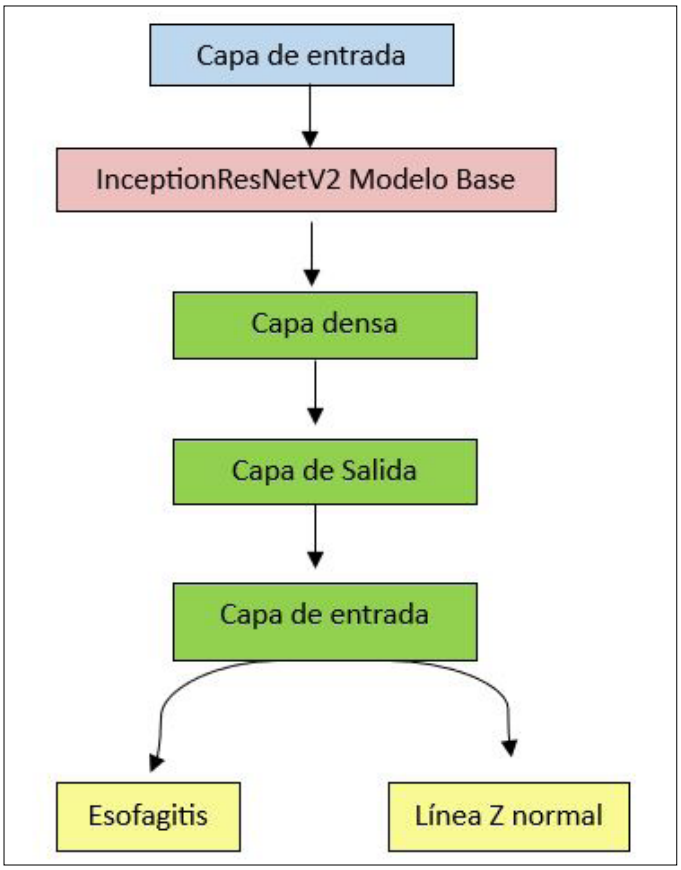


***Figura 1.*** ***Esquema del modelo creado, en verde las capas añadidas al modelo base InceptionResNetV2 que se muestra en rojo.***

El proceso de ajuste fino (fine-tuning en inglés) se llevó a cabo manteniendo fijas las capas iniciales del modelo –las responsables de extraer características– y reentrenando las capas finales específicas. De esta forma, aprovechamos la riqueza de los pesos obtenidos de ImageNet y orientamos la red hacia la discriminación de nuestras dos categorías clínicas de interés. Esto permite un uso mucho más eficiente de los datos y del tiempo de entrenamiento, al evitar entrenar el modelo desde cero.

Entrenamos el modelo con el conjunto de imágenes endoscópicas KVASIR[6], un repositorio que abarca imágenes de endoscopia digestiva. En particular, utilizamos 2000 imágenes correspondientes tanto a la línea Z normal como a diferentes grados de esofagitis, dividiendo los datos en un 80% para entrenamiento y un 20% para validación. Este balance en la partición de los datos permitió una formación robusta del modelo y una evaluación preliminar fiable de su desempeño.

Para llevar a cabo una validación más exhaustiva, empleamos el conjunto de imágenes HyperKVASIR[7], que suma en total 1164; 932 de línea Z normal y 232 de esofagitis. En esta fase, se excluyeron los casos más leves, centrándonos únicamente en los grados B, C y D de la clasificación de Los Ángeles, con el fin de evaluar la capacidad del modelo para identificar lesiones más avanzadas. Adicionalmente, se incorporó un tercer conjunto de imágenes procedentes del Hospital Regional Universitario de Málaga. Este conjunto constaba de 203 imágenes de esofagitis (todas ellas patológicas) que incluían distintas gradaciones de severidad (76 imágenes de grado A de los Ángeles, 42 de grado B, 28 del grado C, 22 del grado D y 18 en la categoría de otros destinada a cuando el endoscopista no especificaba el grado de esofagitis) reforzando así la diversidad y la representatividad clínica de los datos utilizados en el estudio (Tabla 1).

## Resultados

A continuación, se presenta un ejemplo detallado de la capacidad de predicción individual de nuestro modelo, mostrando los porcentajes de confianza asignados a cada categoría diagnóstica. Para ilustrar este aspecto, hemos seleccionado de manera aleatoria cinco imágenes (Figura 2).

Al evaluar la detección en los conjuntos de imágenes, se observó lo siguiente. En el conjunto KVASIR—correspondiente al 20% de imágenes reservado para evaluación y no utilizado durante el entrenamiento—de las 200 imágenes que correspondían a línea Z normal, el modelo identificó correctamente 164, mientras que 36 fueron clasificadas erróneamente como esofagitis. De igual forma, de las 200 imágenes que realmente correspondían a esofagitis, 153 fueron correctamente clasificadas, y 47 se confundieron con línea Z normal (Tabla 2).

Por otro lado, en el conjunto de imágenes HyperKVASIR, la matriz de confusión reveló que, de 932 imágenes de línea Z normal, el modelo clasificó correctamente 833 y falló en 99 casos. Por otro lado, de las 232 imágenes correspondientes a esofagitis, 216 se identificaron correctamente, mientras que


DOI: 10.37352/2025482.1

| Conjunto de imágenes | Cantidad de imágenes | Descripción | Uso |
|---|---|---|---|
| KVASIR[6] | 2000 | Imágenes de línea Z normal y diferentes grados de esofagitis. | Entrenamiento y validación del modelo. |
| HyperKVASIR[7] | 1164 | Imágenes de línea Z normal (932) y esofagitis (232) | Evaluación del rendimiento del modelo en lesiones avanzadas. |
| Hospital Regional Universitario de Málaga | 203 | Imágenes de esofagitis con distintas gradaciones de severidad. | Validación externa en un entorno clínico real. |

***Tabla 1.*** ***Tabla donde se resumen los conjuntos de datos usados en el proyecto.***

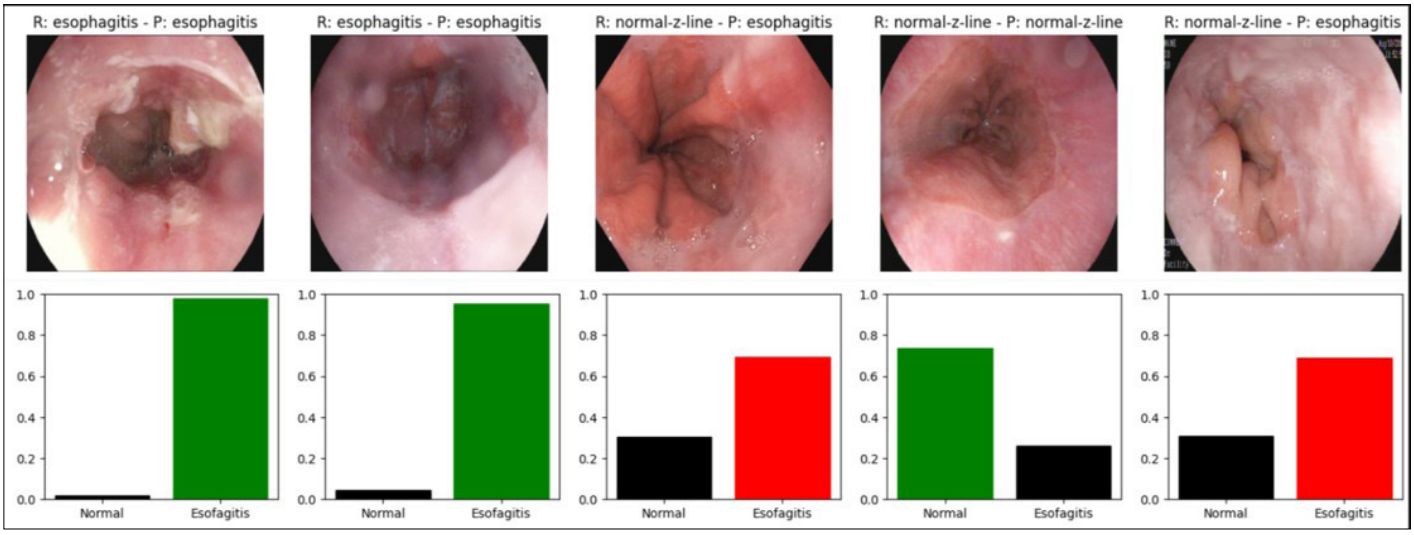


***Figura 2.*** ***Representación de las predicciones individuales que realiza nuestro modelo. Arriba de cada imágen se muestra la etiqueta real (R:) y la etiqueta predicha (P:) y abajo un gráfico de barras con el porcentaje de confianza que asigna a cada clase en la predicción la cual se tiñe de verde si acierta y rojo si falla.***

| KVASIR | Línea Z Normal (Predicho) | Esofagitis (Predicho) |
|---|---|---|
| Línea Z Normal (Real) | 164 | 36 |
| Esofagitis (Real) | 47 | 153 |

***Tabla 2.*** ***Matriz de confusión de la predicción de las 400 imágenes del conjunto de imágens KVASIR que corresponden al 20% de imágenes que hemos reservado para la evaluación.***

| HyperKVASIR | Línea Z Normal (Predicho) | Esofagitis (Predicho) |
|---|---|---|
| Línea Z Normal (Real) | 833 | 99 |
| Esofagitis (Real) | 16 | 216 |

***Tabla 3.*** ***Matriz de confusión de la predicción de las 400 imágenes del conjunto de imágens HyperKVASIR.***

16 fueron erróneamente catalogadas como línea Z normal (Tabla 3).

Para comprarar ambas evaluaciones usaremos la métrica de la curva ROC, que es esencial para determinar el desempeño global de nuestro sistema de clasificación. Un aspecto relevante fue la exclusión de esofagitis grado A únicamente en la base de datos HyperKVASIR, con el fin de evidenciar que al contar con un mayor contraste entre la imagen de control y la imagen patológica, el modelo puede discriminar con mayor eficacia, lo que favoreció la eficiencia en la detección de la patología, logrando valores de área bajo la curva (AUC) de 0.884 para el conjunto de datos KVASIR y 0.970 para el conjunto de datos HyperKVASIR. Estos resultados apuntan a un alto nivel de exactitud diagnóstica (Figura 3).

Asimismo, el gráfico siguiente ilustra el porcentaje de aciertos obtenidos por el modelo al ser evaluado con la cohorte del Hospital Regional de Málaga. Este análisis independiente es fundamental para corroborar la aplicabilidad del modelo en entornos clínicos diversos y validar la solidez de la metodología en circunstancias reales (Figura 4).

Por último, al clasificar el rendimiento según la severidad de la esofagitis, se observó una correlación positiva entre el incremento en la severidad y el porcentaje de aciertos. Este hallazgo sugiere que el algoritmo resulta particularmente eficiente al detectar lesiones más avanzadas (Figura 5).


DOI: 10.37352/2025482.1

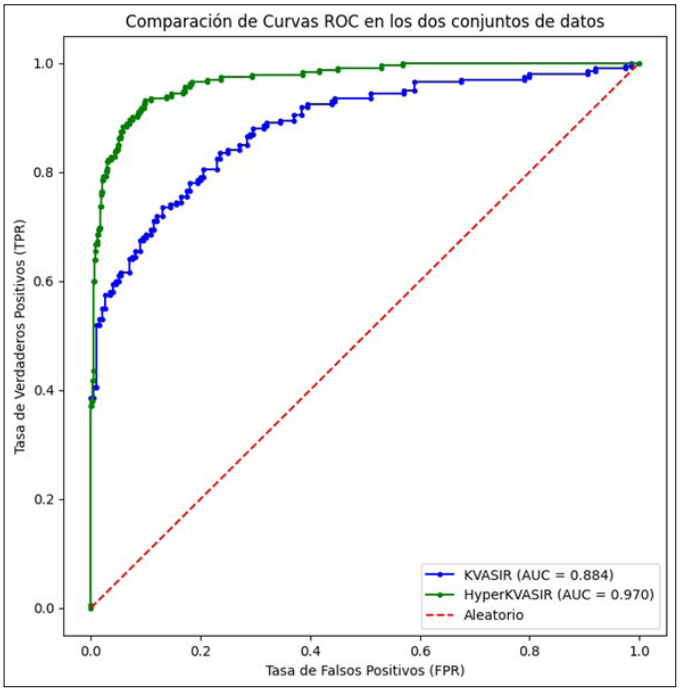


*Figura 3.* ***Representación de las curvas ROC en la validación con el KVASIR (que incluye esofagitis de todo grado de severidad) y HyperKVASIR que solo incluye grados B, C y D descartando los casos más leves) en el que se observa una mejor curva en este último conjunto.***

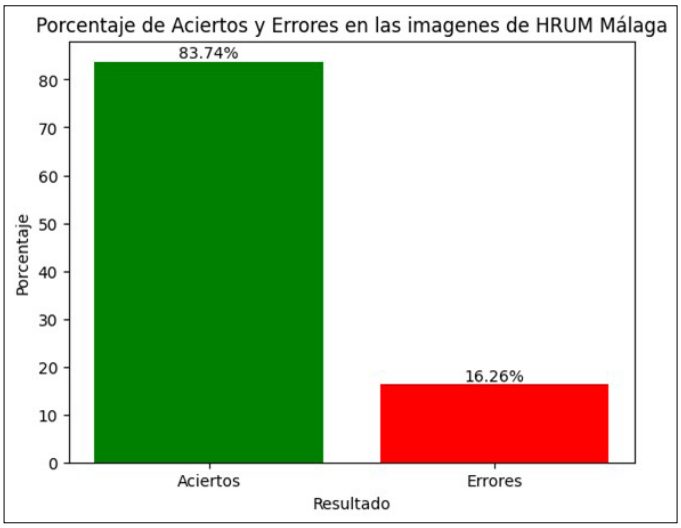


*Figura 4.* ***Porcentaje de aciertos en las imágenes del conjunto creado con casos del Hospital Regional Universitario de Málaga.***

## Discusión

Los hallazgos de este estudio ponen de manifiesto el potencial de la arquitectura InceptionResNetV2 para la detección y clasificación de esofagitis en base a imágenes endoscópicas. El uso de capas preentrenadas con el extenso conjunto de datos ImageNet, junto con el ajuste fino enfocado en el problema de esofagitis vs. línea Z normal.

El valor de la AUC (área bajo la curva ROC) obtenido en los dos principales conjuntos de validación —KVASIR e

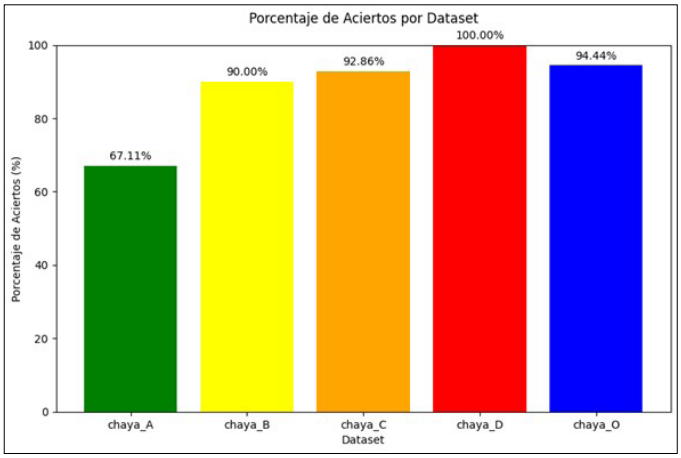


*Figura 5.* ***Porcentaje de aciertos estratificado por el grado de severidad de la clasificación de los ángeles (grado A, B, C y D) la categoría chaya_O (Otros) incluye todas las imágenes en las que el endoscopista no especificó el grado.***

HyperKVASIR— permitió evidenciar tanto la consistencia como la capacidad de generalización del modelo. La exclusión de grados más leves de esofagitis (grado A) en HyperKVASIR mostró cómo un mayor contraste entre las imágenes normales y las patológicas facilita una discriminación más nítida, reforzando la hipótesis de que el modelo funciona de forma especialmente robusta en lesiones más severas. Este aspecto adquiere relevancia clínica, dado que, en la práctica, las lesiones avanzadas suelen requerir un diagnóstico y tratamiento más oportunos.

El análisis independiente en el conjunto de imágenes del Hospital Regional Universitario de Málaga aporta evidencia adicional sobre la aplicabilidad de la propuesta en entornos diversos. Los resultados confirman la utilidad de la metodología no solo en bases de datos públicas, sino también en un contexto clínico real, con variaciones en las condiciones de captura de la imagen, tipos de equipos endoscópicos y características poblacionales.

El hecho de que el rendimiento del modelo aumente en relación con el grado de severidad de la esofagitis sugiere que la red neuronal es capaz de detectar con mayor precisión las alteraciones estructurales más evidentes. Sin embargo, se hace necesario profundizar en la clasificación de lesiones incipientes, ya que una identificación temprana resulta esencial en la práctica médica para prevenir complicaciones futuras y mejorar el pronóstico de la condición.

A pesar de los resultados prometedores, este estudio presenta algunas limitaciones. Por un lado, el número total de imágenes, aunque significativo, podría ampliarse para abarcar una mayor representatividad de las diferentes formas de presentación de la esofagitis, en especial las de grado A.


DOI: 10.37352/2025482.1

Por otro lado, factores como la variabilidad en la calidad de la imagen y la presencia de artefactos durante la endoscopia pueden influir en la exactitud del modelo.

## Conclusiones

Aunque la utilidad de este modelo resulta todavía poco aplicable a la práctica clínica, este estudio pone en evidencia el potencial de las redes neuronales profundas en la endoscopia y subraya la importancia de la colaboración entre hospitales para crear bases de datos multicéntricas. Aumentar tanto la cantidad como la diversidad de imágenes es crucial para entrenar modelos que, en el futuro, puedan implementarse en contextos clínicos de mayor relevancia. Los resultados aquí obtenidos resaltan que, si bien el modelo muestra alta sensibilidad en lesiones avanzadas, aún se enfrentan retos en la identificación de estadios incipientes y en la adaptación a diversas condiciones de captura de imágenes. Con vistas a la práctica, la validez multicéntrica, el uso de técnicas de aumento de datos y la integración de estos sistemas en flujos de trabajo clínicos podrían, a largo plazo, favorecer diagnósticos más ágiles y precisos para una variedad de patologías gastrointestinales.

## Bibliografía


1. *Ripoll C, Groszmann R, Garcia–Tsao G, Grace N, Burroughs A,*
1. *Okagawa Y, Abe S, Yamada M, Oda I, Saito Y. Artificial Intelligence in Endoscopy. Dig Dis Sci 2022;67(5):1553-72.*

2. *Namikawa K, Hirasawa T, Yoshio T, Fujisaki J, Ozawa T, Ishihara S, et al. Utilizing artificial intelligence in endoscopy: a clinician's guide. Expert Rev Gastroenterol Hepatol 2020;14(8):689-706.*

3. *Pang B, Nijkamp E, Wu YN. Deep Learning With TensorFlow: A Review. J Educ Behav Stat 2019;45(2):227-248.*

4. *Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning [Internet]. arXiv [Preprint] 2016. Available from: URL: https://arxiv.org/abs/1602.07261*

5. *Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2009 Jun; Miami, Florida, USA. p. 248-255. doi:10.1109/CVPR.2009.5206848. Available from: https://ieeexplore.ieee.org/abstract/document/5206848.*

6. *Pogorelov K, Randel KR, Griwodz C, Eskeland SL, de Lange T, Johansen D, et al. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection [Internet]. Association for Computing Machinery 2017. Available from: URL: https://doi.org/10.1145/3193289.*

7. *Borgli H, Thambawita V, Smedsrud PH, Hicks S, Jha D, Eskeland SL, et al. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Sci Data 2020;7(1):283.*



DOI: 10.37352/2025482.1